Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR036686

Donghui Li and Yanan Chen are co-senior authors.

Key Points:

- Five basic types of operation modules are categorized for 452 reservoirs
- Seasonal patterns for module application transition are identified for reservoirs with different sizes, operation purposes, and locations
- The basic types of modules and operation patterns inform reservoir operation modeling

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
X. Cai,
xmcai@illinois.edu

Citation:
Li, D., Chen, Y., Lyu, L., & Cai, X. (2024). Uncovering historical reservoir operation rules and patterns: Insights from 452 large reservoirs in the contiguous United States. *Water Resources Research*, *60*, e2023WR036686. https://doi.org/10.1029/2023WR036686

Received 17 NOV 2023
Accepted 24 JUL 2024

Uncovering Historical Reservoir Operation Rules and Patterns: Insights From 452 Large Reservoirs in the Contiguous United States

Donghui Li[1], Yanan Chen[1], Lingqi Lyu[1], and Ximing Cai[1]

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Abstract Reservoir operations are influenced by hydroclimatic variability, reservoir characteristics (i.e., size and purpose), policy regulation, as well as operators' experiences and justification. Data-driven reservoir operation models based on long-term historical records shed light on understanding reservoir operation rules and patterns. This study applies generic data-driven reservoir operation models (GDROMs) developed for 452 data-rich reservoirs with diversified operation purposes across the CONUS to explore typical operation rules and patterns. We find that the operating policies of any of these reservoirs can be modeled with a small number (1–8) of typical operation modules. The derived modules applied to different conditions of the 452 reservoirs can be categorized into five basic types, that is, constant release, inflow-driven piecewise constant release, inflow-driven linear release, storage-driven piecewise constant release, and storage-driven nonlinear (or piecewise linear) release. Additionally, a joint-driven release module, constructed from these five basic types, has been identified. The analysis further shows the module application transition patterns featuring operation dynamics for reservoirs of different operation purposes, sizes, and locations. The typical module types can be used as "Lego" bricks to build operation models, especially for data-scarce reservoirs. These module types and their application and transition conditions can inform Standard Operation Policy (SOP) and Hedging Policy (HP) with specific inflow, storage, and/or both conditions.

1. Introduction

Dams and reservoirs have been extensively constructed and operated to serve human societies around the world. Globally, there are over 50,000 large dams (higher than 15 m), and the cumulative storage capacity is approximately 8,000 km³. This volume is comparable to 15% of the annual river flow into the oceans (Hanasaki et al., 2006; Lehner et al., 2011). The United States is a notable "dam nation," with all major rivers being dammed and a total storage capacity that approximates the mean annual runoff (Graf, 1999). Thus, reservoirs play a critical role in human societies by regulating streamflow for water supply, irrigation, flood control, hydropower generation, navigation, and recreation. While numerous efforts have been made to simulate or/and optimize reservoir operations for individual reservoirs via either physical-based or data-driven models, this paper explores typical operation modules and their application transition patterns for a broad range of reservoirs with diverse reservoir characteristics like size, operation purpose, and geographic location (climate). We hypothesize that those typical modules functionally respond to various inflow and storage conditions and release and/or storage demands, for example, storage for future water supply, recreation, or regulating peak flow for flooding control; release for navigation, water supply, and/or irrigation; both release and storage for hydropower generation, etc.

Despite longstanding efforts for reservoir operation studies (Labadie, 2004; Wurbs, 1993; Yeh, 1985), our modeling capability and understanding of real-world reservoir operation complexities remain limited. One of the main reasons why knowledge gap exists is intractable human behaviors of real-world reservoir operators. While operators must adhere to regulations such as operation curves, they usually have space to use their own judgment and experience in real-world practices when facing uncertainties in inflow and water demands (Oliveira & Loucks, 1997), especially during an extreme event (e.g., drought or flood). Such behaviors of reservoir operations are usually intractable, presenting significant challenges for understanding and modeling reservoir operations in the real world.

To fill this knowledge gap, many recent studies attempted to extract reservoir operation behaviors from historical records. Among those studies, some employed supervised machine learning techniques to replicate observed release using predictor variables for selected case study reservoirs (e.g., Coerver et al., 2018; Corani et al., 2009;

Ehsani et al., 2016; T. Yang et al., 2016, 2020). For example, T. Yang et al. (2016) developed a decision tree model that accurately simulates the operations of several major reservoirs in California. Their model incorporates a myriad of predictors, including reservoir states, hydrological conditions, river indices, as well as the interactions with the California State Water Project. Such heavy data requirements make it difficult to apply the data-driven model to reservoirs with limited data availability. Meanwhile, other studies focused on extracting general operation patterns from historical operations for reservoirs within a region, given growing availability of historical operation data for many reservoirs (e.g., Chen et al., 2022; Giuliani & Herman, 2018; Hejazi et al., 2008; Q. Zhao & Cai, 2020). Following this line of thought, *generic data-driven reservoir operation models* have been developed and validated for hundreds of large reservoirs across the CONUS (Chen et al., 2022; Turner et al., 2020, 2021; Yassin et al., 2019; Q. Zhao & Cai, 2020). Generic models do not depend on reservoir-specific characteristics and are based on operation rules that are applicable to any reservoir; thus, these models can be used to derive generic operation rules for reservoirs when a sufficiently long operation record is available. For example, Q. Zhao and Cai (2020) developed a hidden Markov-decision tree (HM-DT) model to derive representative operation rules for 61 reservoirs in the Upper Colorado River Basin and found that a small number of modules present operation rules under specific operation conditions such as refill or pre-flood release. Following Q. Zhao and Cai (2020), Chen et al. (2022) developed Generic Data-driven Reservoir Operation Models (GDROMs) for 450+ reservoirs in the CONUS. GDROMs identify representative operation modules and their application transition conditions from long-term operation records. It is reasonable to assume that the operation modules derived from such a large number of reservoirs with different sizes, operation purposes, and geographic locations can be used to synthesize some typical or representative operation modules, which might be applicable to any reservoirs and enable more generalized understanding of real-world operations. To this end, we will address the first specific research question of this study: What typical operation modules, especially those that are not well presented in the literature, exist with all the 452 reservoirs?

Reservoir operation dynamics is characterized with seasonal transitions of operation rules, which deal with uncertainties and fluctuations in inflows and water demands. Data-driven models handle the seasonality issue of reservoir operation in different ways. With supervised machine learning methods, the seasonal information is often included implicitly, reflected in input variables like long-term inflow and storage records, and/or embedded in model structures, like the weighted links in neural networks (Ehsani et al., 2016) or the ensemble trees in random forests (T. Yang et al., 2016). Yet, the patterns of these transitions are hidden in a "black box" model and are not explicitly represented. Contrastingly, data-mining studies attempted to shed light on these seasonal transitions directly from observed data (e.g., Giuliani & Herman, 2018; Hejazi & Cai, 2009). Giuliani and Herman (2018) deployed eigen-behavior analysis and clustering techniques on long-term monthly storage data from 172 Californian reservoirs, and identified four typical operation behaviors, each characterized by unique drawdown and refill patterns. In a more recent study, Brunner and Naveau (2023) reconstructed seasonal variations of reservoir operation from streamflow series and analyzed the spatial patterns of reservoir operations in the Alpine region, Europe using daily streamflow data for 10 or more years before and after dam operation to infer reservoir seasonality. In the current study, we will investigate the seasonality of reservoir operations with greater details for 452 reservoirs across the CONUS using daily inflow, storage, and release data with a historical record of 15 years or longer to analyze their seasonal applications and transitions of data-derived operation modules. Specifically, we address the second research question of this study: What are the seasonal patterns for the application transition of typical operation modules for different reservoirs, and how are those patterns associated with reservoir size, operation purpose, and location across the CONUS?

Furthermore, hydrological information is important for reservoir operators to prioritize or deal with various, uncertain hydroclimatic conditions. Gaining clarity on the role of the priority in operators' decisions is critical to understanding realistic operation rules, as well as building more realistic operation models. Hejazi et al. (2008) assessed the importance of a wide range of input variables for daily release decisions across 79 reservoirs in California and the Great Plains. They related the most important variables to specific reservoir attributes, such as size, location, and operation seasons. These identified variables and relationships were then used as potential state variables to develop a more realistic operation optimization model (Hejazi & Cai, 2011). In the current study, we will extend relevant previous studies to a greater number of reservoirs and address the third research question of this study: What hydroclimatic information is prioritized for the operation of different reservoirs?

The three research questions outlined above focus on some crucial aspects of real-world reservoir operations: typical operation modules, seasonal operation patterns, and the role of hydrological information. We will address these questions by employing an approach that is designed not only to synthesize and interpret real-world reservoir operation practices but also to offer a potential framework and insights for building more realistic reservoir models using typical operation modules derived from historical records. In the rest of this paper, we will first provide background on the GDROMs for 452 reservoirs in the CONUS. Following that, we will present the results synthesized from the GDROMs, especially the typical operation modules, module application transition patterns, and prioritized hydroclimatic information in release decisions, in a form that explicitly relates release decisions to water availability conditions (inflow and storage) and reservoir characteristics (size, operation purpose, and location/climate). Furthermore, we will discuss how the typical modules and their application transition patterns can be used to inform the development of more realistic reservoir operation models.

2. Background

The GDROMs for 452 reservoirs are introduced, including the fundamental concepts and methods, data sources and data pre-processing procedures, and model outputs (Li et al., 2023).

2.1. Generic Data-Driven Reservoir Operation Model (GDROM)

The GDROM comprises a small number of representative operation modules derived from the historical operation data for a particular reservoir by the hidden Markov model. An operation module represents a specific operation rule to simulate daily release under specific water availability conditions, such as inflow and storage in a wet/dry period; in the GDROM which is in the form of DT, a module is represented as a branch of the DT (i.e., a sub-DT). The module application conditions are identified by the Classification and Regression Tree (CART, Breiman et al., 1984) algorithm (see Figure S1 in Supporting Information S1 for the GDROM model structure). Eventually, a number of operation modules and their application conditions form the GDROM, with four input variables: daily inflow to the reservoir, initial daily storage, day of the year (DOY, representing seasonality; Bessler et al., 2003), and the Palmer Drought Severity Index (PDSI, represent the climate/weather; Palmer, 1965). These input variables represent the hydroclimatic conditions that trigger different operation modules to simulate release that reflects water demand levels, operational policies, as well as operators' behaviors. For detailed explanation of the GDROM, readers should be referred to Chen et al. (2022) and Q. Zhao and Cai (2020).

Compared with existing data-driven reservoir operation models having complicated model structures and requiring reservoir-specific data, for example, the random forest (RF) model (T. Yang et al., 2016) and the neural network (NN) model (S. Yang et al., 2019), the GDROMs have relatively simple and consistent model structure and common inputs (i.e., inflow, storage, DOY, and PDSI). Although the simulation accuracy with some reservoirs is lower than that from the RF and NN models, the GDROMs gain higher generality and can be applied to most reservoirs with data available from multiple data sets (Chen et al., 2022).

Another merit of the GDROMs is the model interpretability and transparency attributed to the decision tree-based model structure. In particular, the GDROMs depict the operation dynamics, that is, module transition over periods with specified triggering conditions. To extract the time-varied operation patterns, we obtain long-term operation records with high granularity, that is, daily release and storage change, along with hydrological conditions provided by the USBR (2021), the USACE (2021), and the ResOpsUS database (Steyeart et al., 2022a, 2022b). In addition, we retrieve monthly PDSI data statewide from the portal of the National Oceanic and Atmospheric Administration (NOAA; 2021).

Moreover, the GDROM has been shown to capture long-term operation rule changes represented by the identified emerged hidden Markov states, which reflect changed operation rules over years. Within the GDROM, operation modules are re-sequenced to model the inter-year changes, that is, different modules are applied during the same season in earlier and more recent years (Chen et al., 2022; Q. Zhao & Cai, 2020). These operational changes are associated with both altered hydroclimatic conditions (e.g., inflow), water demand (e.g., release), and/or policy changes, which are reflected in the input variables, and in turn captured via CART training and validation. A detailed description of training the GDROM for reservoirs exhibiting intra- and inter-year operational changes is referred to Chen et al. (2022).

We name DGROMs as generic models because GDROMs applied to any of the 452 reservoirs have the common inputs and they are structured with common module types (Section 3.2). In addition, they are derived using a common data mining method relying on historical records to derive operation rules, therefore capturing implicitly

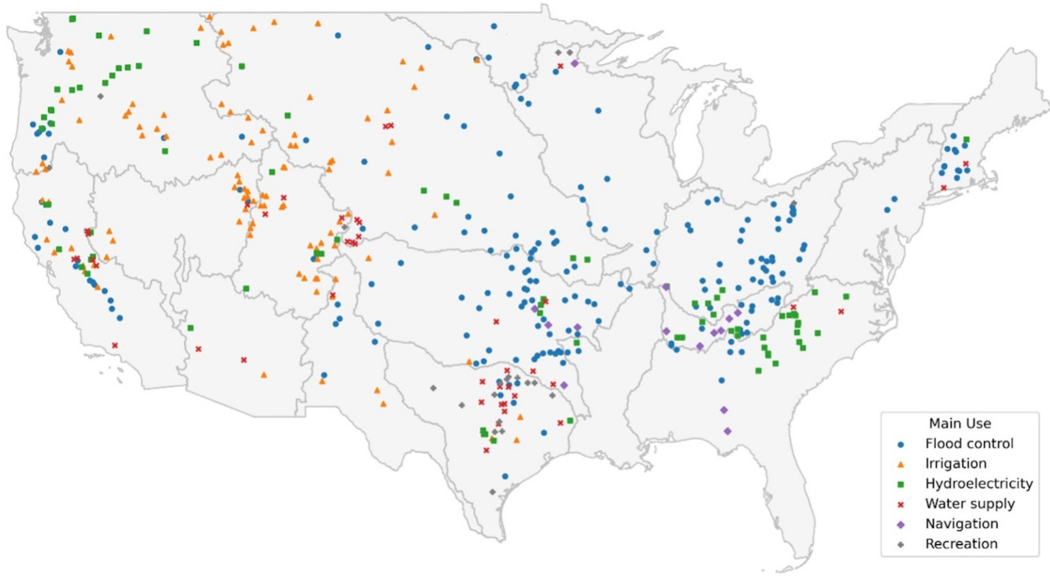

Figure 1. Location of 452 reservoirs across the CONUS with their primary operation purposes, retrieved from the GRanD database (Lehner et al., 2011).

specific reservoir characteristics such as reservoir size, seasonality, and water demand. Thus, they do not use reservoir characteristics as direct inputs.

2.2. Overview of 452 Reservoirs in the CONUS With Long-Term Operation Records

The GDROM has been applied to 452 large reservoirs across the CONUS (Figure 1), which regulate the major rivers in the United States for multiple purposes, including flood control, water supply for irrigation and municipal demands, and hydropower. Many reservoirs are operated with seasonally specific rules, manifested by intra-year variation of storage and release, in response to seasonally varied hydro-climatic conditions and water demands. Some of the large reservoirs regulate the river system inter-annually and play a vital role in water conservation for drought mitigation. Notably, the operation policies of some reservoirs have changed, or will likely change in the future in response to non-stationary hydrological conditions, changed water demand, and more restricted environmental requirements (Conway & Mahé, 2009; Giuliani & Herman, 2018; Wallington & Cai, 2020). Among the 452 reservoirs, those primarily operated for flood control take the largest portion (43%), which are primarily located in Eastern and Central United States; following flooding control is irrigation (23%), mostly distributed in the Western United States. We also have hydropower reservoirs (17%) primarily located in the Southeastern United States and the Pacific Northwest, water supply reservoirs (9%), recreation reservoirs (5%), and navigation reservoirs (3%) in the various CONUS regions. The majority length of the records is 15+ years, most of which are sufficiently long to contain inter-annual operation patterns and long-term changes.

The data preprocessing for the 452 reservoirs follows our recently published work (Chen et al., 2022). In real-world reservoir operations, it is difficult to measure the actual inflow to a reservoir with accuracy because of evaporation loss, seepage loss, and recharge/discharge between a reservoir and the surrounding aquifer (Deng et al., 2015). The net inflow computed based on water balance is often used (USBR, 2021). Compared to inflow, the release and storage observations bear fewer errors (i.e., release is human-controlled; storage can be converted from easily observed elevation), and the errors involved in the release and storage data obtained from USBR or USACE are ignored in this study. This study employs observed release and storage to calculate net inflow, as detailed in Chen et al. (2022).

The raw data are also processed for the requirement of training the GDROMs. Since the observed operation (i.e., daily release series) is assumed to follow the Markov process, the training samples must be continuous, while multiple continuous pieces are also acceptable. Thus, we detect and remove the missing dates and break the operation record into multiple continuous pieces from the missing data points. Note that the pieces without sufficient observations cannot capture the latent temporal dependencies of the release decisions, and thus only pieces with more than 100 continuous observations are retained for training. These segmented continuous time series are treated as independent samples for model training (Q. Zhao & Cai, 2020). In addition, for each reservoir, the inflow, storage, and release are normalized by the maximum historical storage during the observation period, which avoids the effect of reservoir size and reduces the time required for hyperparameter tuning.

In addition, there are some outliers with abnormal sudden storage changes in the operation records, which might be due to measurement errors (or documentation typos). These storage outliers often result in a large negative net inflow followed by a positive net inflow with a similar absolute value in two consecutive days, and vice versa. To resolve this issue, we detect the days with outliers and replace the storage values with linearly interpolated values between the normal storage values in two adjacent days, as detailed in Li et al. (2023).

2.3. Data Sets of 452 Reservoirs Across the CONUS

As demonstrated by Chen et al. (2022), the GDROM can provide reliable release predictions for various reservoir types, as evaluated by the Nash-Sutcliffe Efficiency and Percent of Bias metrics (Moriasi et al., 2007). By applying the GDROM to model 452 large reservoirs throughout the CONUS, we have constructed a comprehensive data set containing data-driven operation rules for these reservoirs and made this data set openly accessible via HydroShare (Li et al., 2023). For each reservoir, the derived operation rules comprise representative operation modules and corresponding module application conditions, which form the basis of the subsequent analyses presented in this study.

The GDROM-based data set for deriving reservoir release rules is not the first inventory at the CONUS level; another notable example is the ISTARF-CONUS developed by Turner et al. (2021). However, considerable differences in the model structures yield distinct capabilities for specific application scenarios. ISTARF constructs a release function using a predetermined structure based on reservoir states (e.g., storage), characteristics (e.g., capacity), and seasonal variability, with function parameters estimated via regression. Importantly, these parameters can be extrapolated to data-scarce reservoirs, enabling ISTARF to parameterize the operation of 1930 large reservoirs in the CONUS. On the other hand, GDROM, a semi-supervised machine learning model leveraging hidden Markov and decision tree techniques, does not predefine the release rule formula or operation model structure. Instead, it automatically identifies operation modules from historical records, providing greater flexibility when modeling various reservoirs (Hastie et al., 2009a). Furthermore, GDROM can depict the operation dynamics and thus facilitates the discovery and comparison of operation patterns, which is the primary focus of this study.

3. Representative Module Categorization and Operation Pattern Analysis

In this section, we first analyze the representative modules of the GDROM for the 452 reservoirs. This is accomplished through a synthesis of all the modules derived for different reservoirs, which are subsequently categorized into six typical types. Following that, we show the seasonal operation patterns in terms of module selection and transition for reservoirs with different primary operation purposes. We further investigate the role of hydrological information in the seasonal operation patterns considering the impact of reservoir size, operation purpose, and location. It should be noted that some results in the following sections are intuitive, which are presented to demonstrate the model capability for extracting real-world operations. Meanwhile we will highlight new insights that expand understanding of reservoir operations in practice.

3.1. The Number of Modules for Different Reservoirs

The modules characterize the release decisions of a reservoir in a certain period (e.g., of flooding control, irrigation, storage refill, or drought management) and respond to a certain hydrological condition in terms of inflow and/or storage. It is found that different numbers of modules are identified for reservoirs with different operation purposes, reflecting the diversity of operation schemes applied in reservoir operation practice. Figure 2a shows the distribution of the number of modules that are identified from long-term historical operation records. A notable observation is that the operation of any of these reservoirs can be effectively approximated using 1–8 modules, irrespective of their size, operation purpose, or location. The majority of these reservoirs require only a minimal number of modules, with most requiring less than six modules. To break it down, 25% of the 452 reservoirs use just one module; 54% use two or three; 19% use four or five; only 2% require six or more.

10.1029/2023WR036686

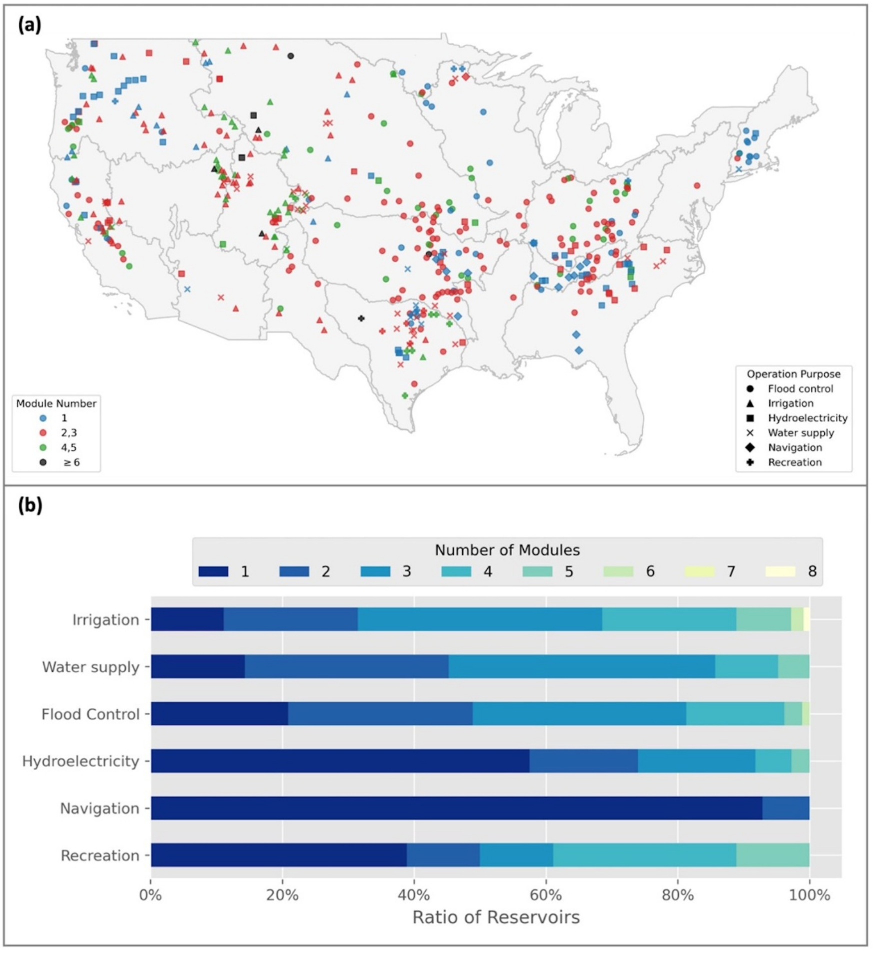

Figure 2. Distribution of number of extracted operation modules, (a) for individual reservoir across the CONUS, (b) for reservoirs grouped by their primary operation purposes.

A notable spatial pattern in the numbers of modules is observed in the Pacific Northwestern, Tennessee, and North Atlantic regions. In these areas, the majority of the reservoirs are found with only a single module. Specifically, 62% of single-module reservoirs in the Pacific Northwestern region are primarily operated for hydroelectric power generation; 91% in the Tennessee region are mainly utilized for either hydropower or navigation. This is also shown in Figure 2b, the module numbers with different operation purposes. In contrast, in the North Atlantic region, the reservoirs are predominantly used for flood control. However, the storage capacity of these reservoirs is relatively small (all falling below 10,000 acre-feet), which sets limited operation function, thereby precluding multi-module operations (see more discussion on flood control reservoirs in Section 3.3.3). In addition, reservoirs operated for navigation in any region in sample have a limited storage capacity, and the release mainly depends on the inflow; therefore, the operation rules for those reservoirs do not change significantly.

Among the reservoirs with different operation purposes and different module numbers shown in Figure 2b, reservoirs primarily operated for hydropower present an interesting case. Nearly 80% of those reservoirs are operated with only a single or two modules, and 20% are operated with three or more modules. This results from mixed types of hydropower reservoirs in the real world: (a) the run-of-river hydropower dams (e.g., John Day Dam on the Columbia River) harvesting energy from streamflow with small storage, which are operated by one or two modules; (b) the impoundment hydroelectric facilities (e.g., Glen Canyon Dam on the Colorado River), with large regulation capacity, which are operated by more than two modules to respond to complex conditions and the transition between the conditions.

Irrigation reservoirs usually operate in a seasonal manner; flood control reservoirs are usually operated with other purposes such as water supply, irrigation, and hydropower; the operation of recreational reservoirs corresponds to inflow, storage, and seasonality. Therefore, the numbers of modules for the reservoirs with irrigation, water supply, flood control, and recreation as a primary operation purpose are relatively large. Specifically, 89% irrigation reservoirs, 79% flood control reservoirs, 86% water supply reservoirs, and 61% recreation reservoirs employ two or more modules.

As expected, the operation of large reservoirs vitally supplying water for populated regions is featured with multiple modules; for example, Lake Powell in the Upper Colorado River Basin regulates streamflow interannually and employs four modules in its operations. Especially, large reservoirs serving multiple purposes usually adopt various operation schemes, necessitating the use of several operation modules. The Fort Peck Dam, for example, the highest dam along the Missouri River, utilizes six operation modules to achieve its multi-objective operation, including flood control, irrigation, and hydropower.

3.2. Classification of Module Categories

Among the 1,155 modules derived from the 452 reservoirs in the CONUS (though only a small number of modules for a particular reservoir), five typical module types are identified with respect to the primary driven variables, namely, constant release, inflow-driven piecewise constant release, inflow-driven linear release, storage-driven piecewise constant release, and storage-driven nonlinear (or piecewise linear) release, as shown in Figure 3. The cumulative distribution of the magnitude of release associated with each type of the modules is illustrated in Figure S2 in Supporting Information S1.

The constant release modules represent the operation scheme characterized by constant or near constant releases, regardless of the inflow and storage levels (see Figure 3a). The constant release modules are mainly found in operations with low releases during a low-flow period, such as the water discharge during the refill season for building reservoir storage or the restricted releases during dry periods for water conservation. For example, the Clark Canyon Dam in Montana State constantly has low releases from November to April for storage building (and high release from May to September, Figure S4 in Supporting Information S1). In addition, the constant release modules are commonly observed with extremely large reservoirs, where the releases are relatively stable and demand-driven, with little effect from inflow and storage. For example, the Garrison Dam in North Dakota State applies two constant release modules in most of the months, and the non-constant release modules are only applied in response to flood events (Figure S5 in Supporting Information S1).

The two types of Inflow-driven modules are demonstrated in Figures 3b and 3c. Under the inflow-driven piecewise constant release module, the release is constant within a certain inflow range. For the inflow-driven linear release modules, the release changes linearly with the inflow. The inflow-driven piecewise constant release modules are often applied for low releases; conversely, the inflow-driven linear release modules are more applied for high releases (Figure S2 in Supporting Information S1). The Yatesville Dam in Kentucky State that is primarily operated for flood control, demonstrates the different applications of the two inflow-driven modules (Figure S6 in Supporting Information S1).

Similar to the inflow-driven modules, the storage-driven modules use reservoir storage as the most decisive factor, and the storage-driven piecewise constant release modules and storage-driven nonlinear release modules are illustrated in Figures 3d and 3e. Under the storage-driven piecewise constant release modules, releases are decided in a piecewise constant manner with the reservoir storage (i.e., constant releases remain at different values corresponding to different storage levels). This type of modules is mainly observed with low or normal releases. These modules usually maintain certain storage levels. By comparison, under the storage-driven nonlinear release modules, the release remains relatively low when the reservoir storage is less than a threshold; beyond the threshold, the release significantly increases. This reflects the hedging policy (You & Cai, 2008) adopted in practice as a conservative strategy during dry periods. The Deerfield Reservoir in South Dakota presents an example (Figure S7 in Supporting Information S1). The storage-driven nonlinear release modules are also applied during the transition between the refill period and the high-release period. For example, the Fresno Reservoir has a storage-driven nonlinear release module applied in April and September, the transition months between low- and high- releases (Figure S8 in Supporting Information S1).

Besides the five basic module types, the joint-driven modules do not determine the release solely depending on inflow or storage but on the combined information of both. The decision trees of the joint-driven modules are characterized by multiple layers with both inflow and storage as splitting nodes. In other words, the conditions of

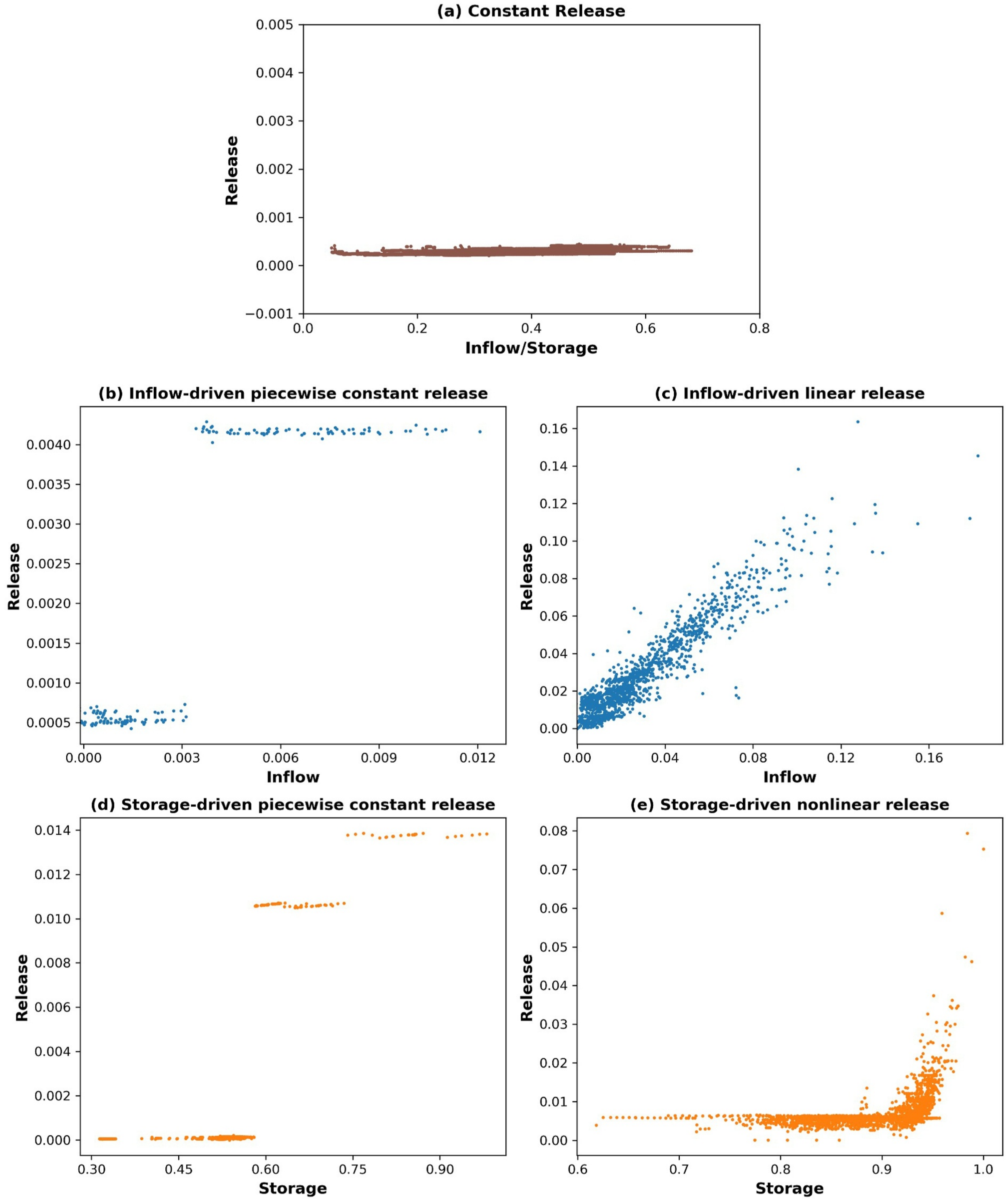

Figure 3. Examples for inflow-driven and storage-driven modules: (a) constant release; (b) inflow-driven piecewise constant release, (c) inflow-driven linear release, (d) storage-driven piecewise constant release, (e) storage-driven nonlinear release. Note that for the purpose of illustration, the inflow, storage, and release shown in the figure are observed values taken from different reservoirs for each module type. These values are normalized by maximum reservoir storage. Additional illustrations from a single reservoir are included in Figure S3 in Supporting Information S1.

the decision rules are collectively constructed with inflow and storage. The joint-driven module is associated with a relatively high release level, which is typically applied for flood control, where the inflow and storage are both important in the release decision. For example, the Tom Jenkins Dam in Ohio (Figure S9 in Supporting

Table 1
Application Frequencies of Each Module Type for Reservoirs With Different Operation Purposes

	Constant	Inflow-driven constant	Inflow-driven linear	Storage-driven constant	Storage-driven nonlinear	Joint
Irrigation	48%	5%	19%	9%	8%	11%
Water supply	34%	18%	19%	10%	6%	13%
Hydropower	21%	8%	36%	16%	7%	12%
Flood control	39%	12%	13%	10%	5%	21%

Information S1) illustrates the application of a joint-driven module for high-release operation during the flood season or a high-flow period.

The application frequencies of the five types of operation modules and the joint-modules by operation purposes are provided Table 1, that is, the ratio of days with the application of a module type to all days of a year. Notably, the constant release module type is applied in over 20% of days in a year for all primary operational purposes listed in the table, showing its prevalence in low-release periods. This is particularly evident for irrigation reservoirs (48%), reflecting their typical operation cycle of refill and release and the need for stable releases during the irrigation season. In contrast, hydropower reservoirs exhibit a high application of the inflow-driven linear release module type, which aligns with a large portion of run-of-river hydropower plants among the selected hydropower reservoirs. Flood control reservoirs have a high application of constant release (39%) and joint-driven release modules (21%), indicating the complexity during the flood control period, that is, pre-release before a flood event, storage filling and release below a threshold for downstream safety during the flood event and possibly spill, release after the flood event, and refill around the end of the flooding season. In addition, the application of storage-driven nonlinear module type has relatively low application frequencies 5%–8% among all types of reservoirs, which shows the non-linear form hedging policy has limited applications in the past practices. This is reasonable since non-linear hedging policy is usually applied to extreme conditions such as seasonal or long-term droughts (Zeng et al., 2021).

However, it is important to note the module application is not only affected by reservoir operation purposes, but other factors, as explored in greater depth in the subsequent sections and summarized in Table 2 appearing at the end of Section 3.

3.3. Seasonal Module Selection and Transition Pattern for Different Types of Reservoirs

Multiple operation rules are usually adopted for large reservoirs and applied to different periods within a year in response to intra-annual variation of hydrological conditions, water demands (e.g., the seasonal agricultural water requirements), and/or extreme events such as droughts and floods. We assume that a typical seasonal operation pattern exists with reservoirs of similar characteristics such as sizes, primary uses, and/or locations (associated with climates); while the patterns differ among reservoirs of considerable dissimilarity. To explore this assumption, we identify the most frequently used operation module for each reservoir during each calendar month throughout the entire historical operation period (of a series of years) for the month. Following that, we group reservoirs according to their primary uses, size ratios (defined as the ratio of storage capacity to average annual inflow (Anghileri et al., 2016)), or regions. The percentage of the reservoirs with the most frequently used module type in a month is calculated within each group of the reservoirs (Figures 4–7). Note that only reservoirs with multiple modules are grouped for analyzing the transition between different modules from 1 month to another, and Figures 4–7 display both the selection of module types and the transition between different types of modules.

3.3.1. Seasonal Operation Patterns for Reservoirs of Different Sizes

As shown in Figure 4, the seasonal module type selection and transition are associated with the reservoir size. Reservoirs with relatively smaller size ratios are likely to employ the inflow-driven linear release modules throughout the year. By comparison, reservoirs with larger size ratios exhibit a more frequent application of constant release. For example, for reservoirs with size ratio between 0.0 and 0.2 (relative to the mean annual inflow), if accounting the modules on a monthly basis, on average, the inflow-driven linear release modules are adopted for 58% of the reservoirs. In contrast, for those reservoirs with size ratio larger than 1, the constant release modules are adopted for 71% of the reservoirs. This is consistent with the regular observations that smaller

Table 2
Summary of Operation Conditions for Each Type of Modules

Module type	Representative release levels	Typical reservoir types	Regional applications
Constant release	Low release during low-flow or conservation periods; stable release from large reservoirs	Irrigation reservoirs during storage refill period (80%) Flood control reservoirs during the non-flood season or non-flood days for storm-driven floods (68%) Water supply reservoirs with stable water demand (53%) Large reservoirs for stable releases	Irrigation reservoirs in Western United States (72%)
Inflow-driven piecewise constant release	Low to medium release, driven by critical inflow levels	Medium and small reservoirs (39% for reservoirs with size ratio below 0.5)	Flood control reservoirs (in non-flood periods) in Pacific Northwest Region (60% sampled) and Eastern United States (53% sampled)
Inflow-driven linear release	Medium to high release during high-flow periods; releases from small reservoirs	Flood control reservoirs during flood season (34%)	Irrigation reservoirs (in high-flow or irrigated periods) in Western United States (51%)
		Irrigation and water supply reservoirs with high water requirements (49%) Run-of-river hydropower dams	Flood control reservoirs (in high-flow periods) in Pacific Northwest (56%) and Southeastern United States (69%)
Storage-driven piecewise constant release	Low to medium release, usually determined for maintaining certain storage levels	Reservoirs for recreation, navigation, hydropower, and flood control to maintain certain storage level	Reservoirs in California Region
Storage-driven nonlinear release	Low or high release, depending on the storage level usually during dry periods or transition periods between low- and high-release	Irrigation, water supply and hydropower reservoirs with medium and large storage	Reservoirs in drought-prone regions
Joint-driven release	Medium to high release, usually applied during high-flow periods	All types of reservoirs	Irrigation reservoirs (in irrigation periods) in Pacific Northwest and Upper Colorado Region (44%) Flood control reservoirs (in flood periods) in Pacific Northwest, Arkansas-White-Red, and Ohio Region (71%)

Note. Empirical sampled percentage is provided for typical reservoir types and regional applications.

reservoirs are more likely to be operated according to inflow conditions, while larger ones are less inflow-driven and more regulated by storage, resulting in more stable and consistent release across months.

Additionally, some reservoirs have a medium storage capacity but small streamflow during most months of the year, resulting in relatively large size ratios, and thus the constant release modules with low release levels are applied for these reservoirs in most months. The existence of these reservoirs with a size ratio greater than 1.0 may cause the abruptly increased application of the constant release modules (Figure 4e). Furthermore, seasonal module transitions occur more frequently with the reservoirs with larger size ratios (as can be seen by comparing Figures 4a–4c with Figures 4d and 4e.) The transition between constant (associated with low-release) and inflow-driven (with high-release) modules indicates the switch between storage refill and water release with large reservoirs.

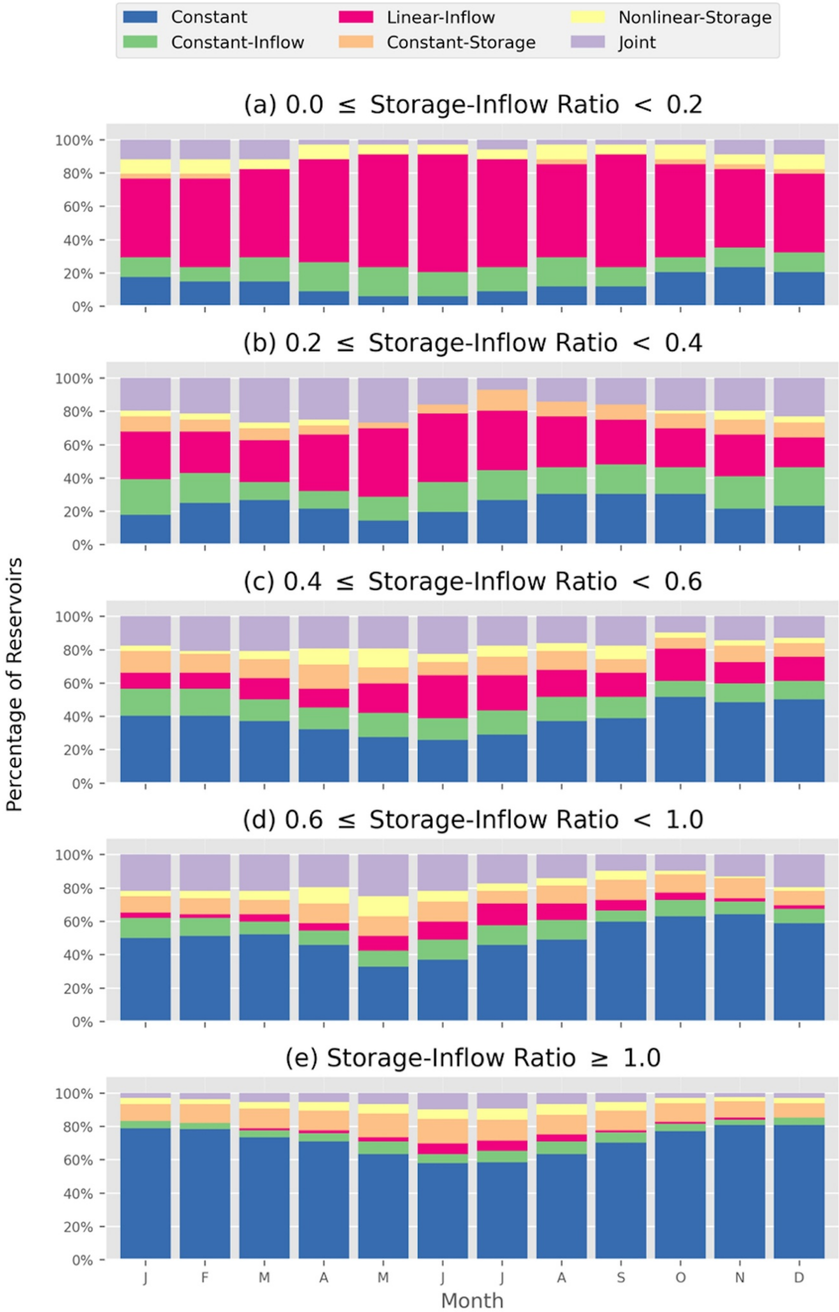

Figure 4. Percentage of reservoirs with monthly most frequently used module type for each size ratio. The vertical axis shows the percentage of reservoirs taking each module type (five basic types and the joint type) as a major module in a given month. Here, constant represents constant release module; linear-inflow represents inflow-driven linear release module; constant-inflow stands for inflow-driven piecewise constant release module; nonlinear-storage represents storage-driven nonlinear release module; constant-storage stands for storage-driven piecewise constant release module; joint stands for joint-driven release module.

3.3.2. Seasonal Operation Patterns for Reservoirs of Different Operation Purposes

Primary operation purposes also play a major role in determining seasonal operation patterns (Figure 5). A notable intra-year transition pattern is observed among the various uses, particularly for irrigation reservoirs, due to the relatively consistent operation for irrigation season, as illustrated in Figure 5a. More specifically, the monthly application frequency of each module type is shown in Table S1 in Supporting Information S1. From

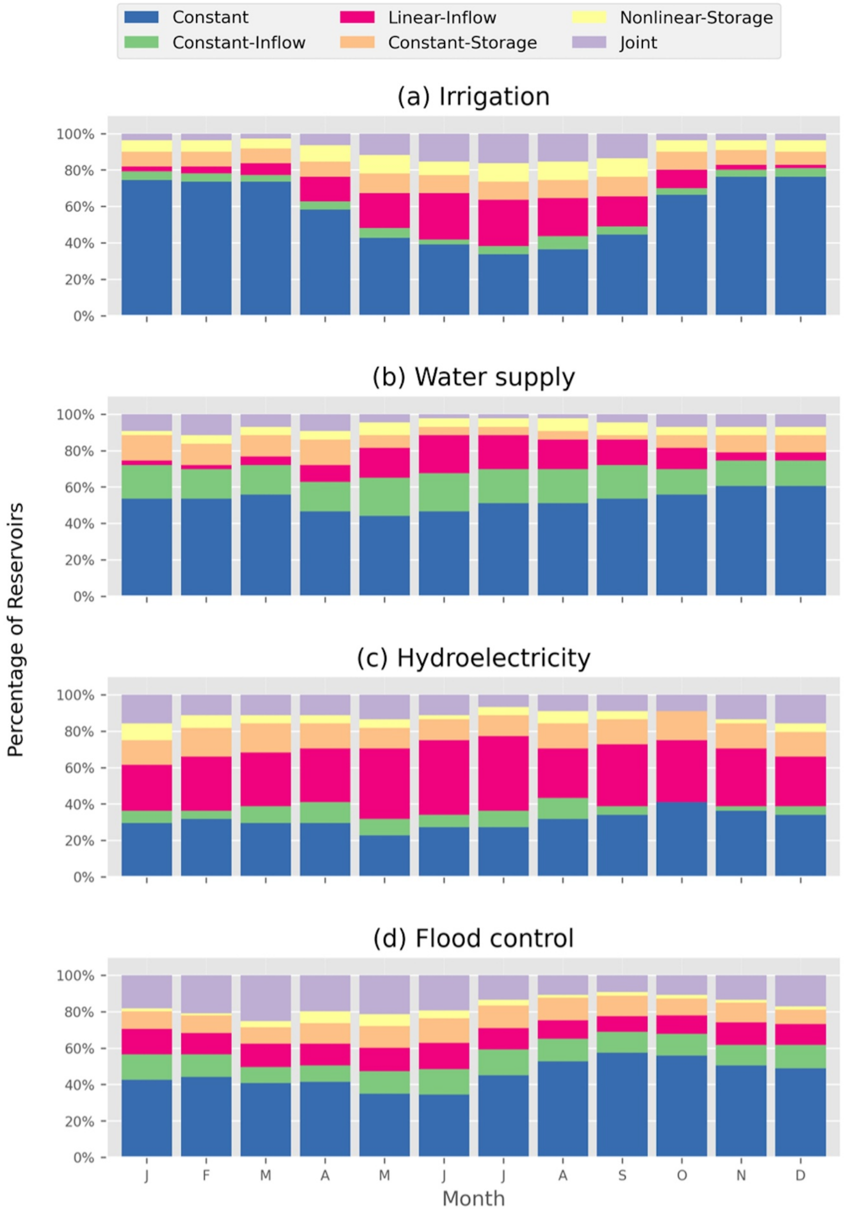

Figure 5. Percentage of reservoirs with monthly most frequently used module type for each primary operation purpose: (a) irrigation, (b) water supply, (c) hydropower, (d) flood control.

September to March, constant release modules prevail for 74% of the irrigation reservoirs on a monthly basis. In contrast, the percentage of constant release modules decreases to 38% during May to July. This seasonal module transition reveals a distinctive seasonal operation cycle of refill and release essential to irrigation reservoirs. Such a cycle is characterized by the shift and transition between (a) the constant release modules utilized for reservoir storage refill during the low-flow season, and (b) water-release modules (e.g., inflow-driven linear release and joint-driven release modules) applied during the irrigation season.

For water supply reservoirs (Figure 5b and Table S2 in Supporting Information S1), the application of the inflow-driven linear release module increases between April and October; meanwhile, the application of the two storage-driven modules decreases. These months lie in the high-flow periods in Texas-Gulf Region, where most water supply reservoirs are obtained for our analysis. The constant release module and the inflow-driven piecewise constant release module have relatively stable applications, which is due to the minor variation of water supply demand across the seasons. For hydropower reservoirs (Figure 5c and Table S3 in Supporting Information S1),

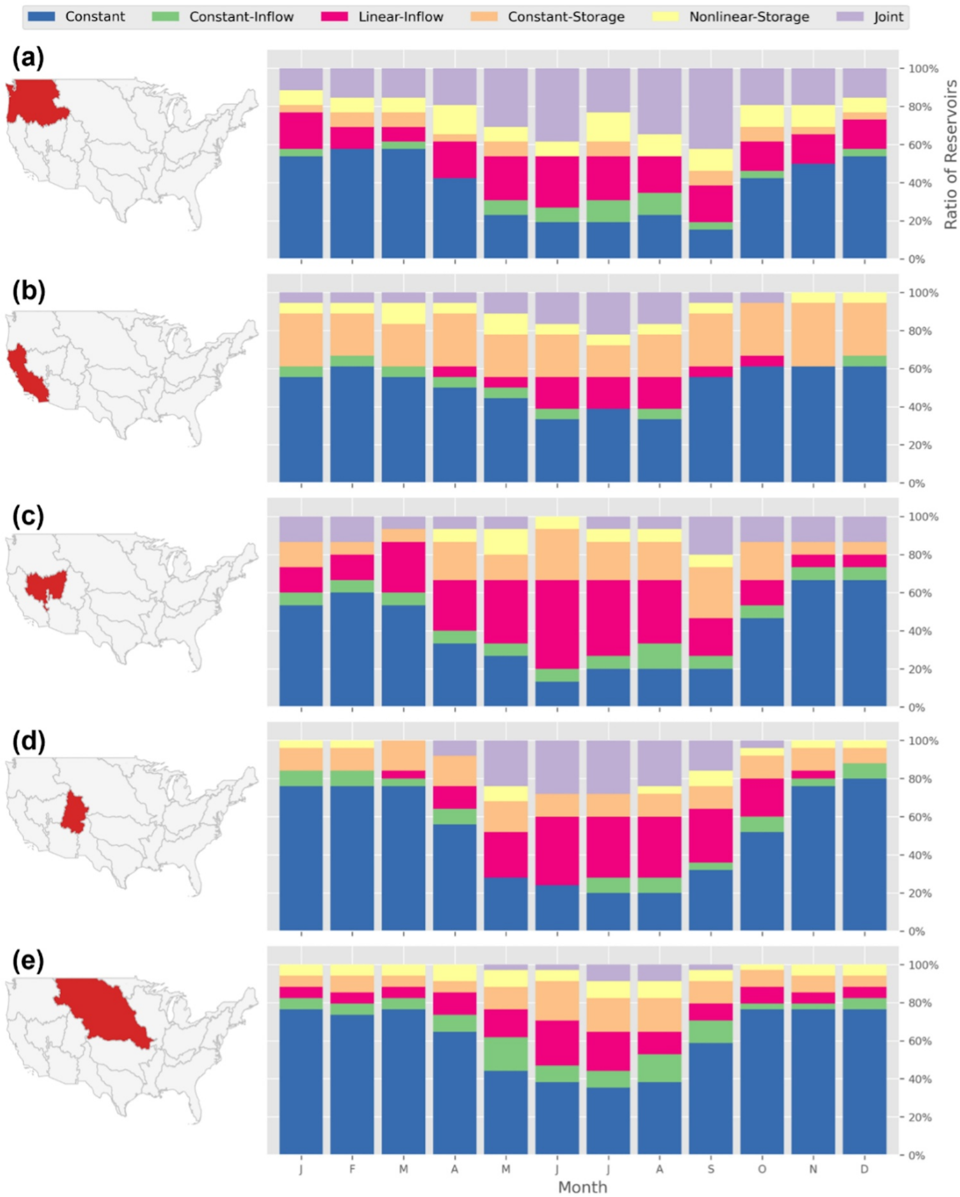

Figure 6. Percentage of irrigation reservoirs with monthly most frequently used module type within selected water regions. The highlighted area in left panel illustrates the specific water region, and the right panel shows the percentage of monthly most frequently used module type for the selected area. (a) Pacific Northwest Region, (b) California Region, (c) Great Basin Region, (d) Upper Colorado Region, and (e) Missouri Region. Note that irrigation is an equally important function of most large reservoirs in the California Region, while it is classified as secondary in the GRanD. For this sake, we also include reservoirs with irrigation as secondary operation purpose in (b).

the relatively simple inflow-driven release rules are mostly adopted throughout a year with occasional transitions to storage-driven or joint-driven operation modules, reflecting the fact that many of the hydropower plants are run-of-river plants. In addition, the stable application of storage-driven piecewise constant release module selection across months implies that a certain level of storage (head) is reserved for hydropower reservoirs underlying storage regulation.

It should be noted that the seasonal module transitions are less significant when reservoirs are grouped by primary uses (Figure 5). As can be observed from Table S4 in Supporting Information S1, the monthly frequency of the six module types are relatively even, especially for inflow- or storage-driven modules. This is because the module

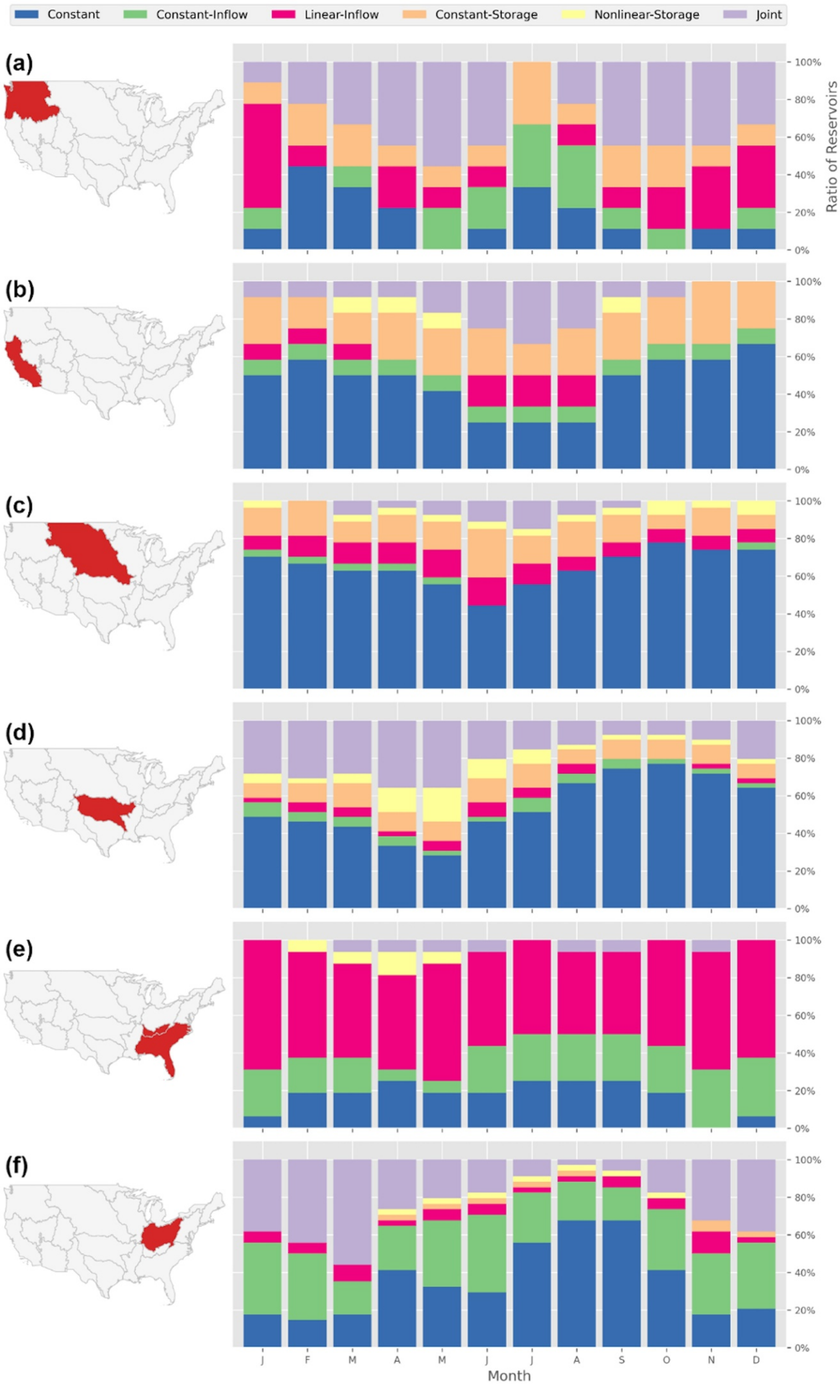

Figure 7. Percentage of flood control reservoirs with monthly most frequently used module type within selected water regions. The highlighted area in left panel illustrates the specific water region, and the right panel shows the percentage of monthly most frequently used module type for the selected area. (a) Pacific Northwest Region, (b) California Region, (c) Missouri Region, (d) Arkansas-White-Red Region, (e) Tennessee Region and South Atlantic-Gulf Region, and (f) Ohio Region.

transition could be hidden when the modules of reservoirs with different flow and water demand conditions are aggregated at the CONUS scale, especially for flood control reservoirs (Figure 5d). The regional specific seasonal transitions are illustrated in the following section.

3.3.3. Seasonal Operation Patterns for Reservoirs at Different Locations

Since hydroclimatic and water use conditions are closely associated with reservoir locations, in the following we discuss the seasonal selection and transition of modules in different regions across the CONUS with irrigation or flood control as the primary use. Given the data availability for this study, we omit the discussion on hydropower since most hydropower plants are run-of-river plants and a relatively simple operation pattern is applied to the reservoirs; we also omit the discussion on water supply reservoirs since they are mainly located in one region (Texas-Gulf Region).

Figure 6 shows the seasonal module selection and transition for reservoirs primarily operated for irrigation in five selected regions in western United States. For all the regions, the inflow-driven linear release module is more applied during the irrigation season but with various percentages of the reservoirs involved among the selected five regions (Figures 6a–6e). Meanwhile, the constant release modules are applied during the non-irrigation season for most of the reservoirs. Specific results are highlighted as follows by region. In Pacific Northwest Region (Figure 6a), besides the irrigation months in the summer, the inflow-driven linear release module is more applied from November to February than in other regions, corresponding to the high rainfall months in this region, where high-release operation rules are employed. In California Region (Figure 6b), the inflow-driven linear release module is less applied in the irrigation month, corresponding to little rainfall in the summertime, and irrigation water supply is mainly provided by the storage built up from the high-precipitation months. Thus, the application of storage-driven modules (storage-driven piecewise constant release module and storage-driven nonlinear release module) is significantly higher in this region than others. Great Basin Region (Figure 6c) shows a higher application of inflow-driven linear release modules during spring, attributed to irrigation requirements during the dry season and the occurrence of high inflow caused by snow melting. Moreover, this region includes a number of small reservoirs (especially in Utah), which need to release more water during the snow-melting months. Upper Colorado Region (Figure 6d) is similar to Pacific Northwest Region in terms of more applications of the joint-driven modules than other regions during the high-flow months, implying that storage is jointly considered with inflow in high-release decisions. In both Missouri Region (Figure 6e) and California Region (Figure 6b), more irrigation reservoirs in the two regions apply the constant release modules and fewer reservoirs apply the inflow-driven linear release modules during the period of May–September, resulting in relatively stable and low releases. This is also related to the common climate feature of the two regions—both are drought-prone and flood events occur very occasionally. In summary, regarding the module selection among regions and crossing months of a year, all regions show a clear transition between the constant release modules during low-flow months and the inflow-driven linear release modules during high-flow months, which represents the common pattern of the seasonal refill-release cycle for irrigation reservoirs.

Figure 7 shows regionally distinctive seasonal operation patterns for the reservoirs primarily operated for flooding control for six selected regions across the CONUS. The module selection and transition show considerable variation across months and among regions. In Figure 7a, the inflow-driven linear release module is more applied from November to January in Pacific Northwest Region than other regions. The joint-driven module is also significantly higher in several months following February, corresponding to the rainy season when flood management is critical in this region. Moreover, there exists a major adoption of the modules with constant releases (i.e., the constant-release, inflow-driven piecewise constant release, and storage-driven piecewise constant release modules) during the summer months (e.g., June–September, the dry season in the region), which are associated with low releases, as observed during this period in Pacific Northwest Region. The gradual transition between higher-release modules (inflow-driven linear release module and joint-driven release module) and lower-release modules (constant-related modules) depicts the distinctive management profiles in wet and dry seasons.

Being opposite to Pacific Northwest Region, in California Region, inflow-driven linear release and joint-driven modules are more applied during summer, and the constant release modules are more applied during winter, as shown in Figure 7b. Since many reservoirs in California Region are managed for water supply and irrigation along with flood control, the storage-driven piecewise constant release module is applied with higher frequency throughout the year, and the transition shown in Figure 7b is affected by the multi-purpose operation of those reservoirs, which is somewhat similar to the irrigation reservoirs in the region (Figure 6b).

For the Missouri Region (Figure 7c), the abrupt switch between low-release and high-release modules, as observed in other regions, is not apparent. This is because most areas in this region are drought-prone and have high agricultural water consumption. Thus, the reservoirs in this region are not operated for flood control during the normal and dry years, and consequently, the seasonality of flood control is hidden, with examples shown in Figures S5 and S10 in Supporting Information S1 for some selected reservoirs. By comparison, for the flood-prone Arkansas-White-Red Region (Figure 6d), it is observed a clear seasonal transition between the flood period and the non-flood period, with joint-driven modules mostly applied for flood control and constant release modules dominantly adopted in the non-flood period. In addition, the seasonal pattern may be underestimated due to the existence of some reservoirs with small storage capacity, which quickly respond to storm-driven floods without exhibiting a significant seasonal pattern, for example, Canton Lake in Oklahoma (Figure S11 in Supporting Information S1).

For the Southeastern United States (Figure 7e), the unique operation patterns show a significantly higher application of inflow-driven modules, especially the inflow-driven linear release module, which agrees with high releases throughout the year in this water-rich region. Figure 7f shows the seasonal module application for flood control reservoirs in Ohio Region, where flood events are mainly episodic and storm-driven, which could occur all year long. Many reservoirs in this region are operated to deal with the various types of flood events, and the typical operation follows the low-release modules (e.g., constant release modules) applied for non-flooding days and the high-release (e.g., joint-driven release modules) modules applied for flood days. The abrupt switch between modules occurs when a flood event starts, as demonstrated in Figures S11–S13 in Supporting Information S1 with three selected reservoirs. Although major floods can occur at any time of the year, those caused by winter storms are more frequent in the region (USACE, 1995). Therefore, it is observed that there is a higher application of high-release modules during winter and a lower application of low-release modules during summer, as shown in Figure 7f.

The summary of seasonal operation patterns for the various types of modules applied to different types of reservoirs are provided in Table 2. Furthermore, the relationship between the regional variability of the patterns and reservoir size is discussed in Text S1 in Supporting Information S1; as shown in Figure S15 in Supporting Information S1, the regional patterns remain for both small and large flood control reservoirs. The potential use of these basic module types and their transition patterns is further discussed in Section 4.

It is well-understood that many reservoirs serve multiple purposes. Here, we illustrate how operation module transitions between different operation purposes can be effectively captured using GDROM decoded modules. We use the John H. Kerr Dam in Virginia as an example, as depicted in Figure 8. The dam is primarily operated for hydroelectric power generation, which requires maintaining reservoir levels within a range to optimize energy production. The reservoir also plays a crucial role in flood control during flooding events. Figure 8 clearly shows that module 0 is predominantly applied to supporting hydropower operations by maintaining a nearly constant storage; meanwhile, module 1 is employed mainly during flooding events, allowing the reservoir storage to increase significantly to help mitigate the risk of peak outflow from the reservoir. The shift from module 0 to module 1 reflects the seasonal operation change from hydropower generation to flood management. During the relatively dry period from 1999 to 2002, high incoming flows with flood risks are minimal compared to normal or wet years. As a result, module 0 are used more frequently, with minimal switch to module 1, indicating that the operations are primarily focused on hydroelectricity.

3.4. The Role of Hydrological Information in the Operation of Various Reservoirs

Hydrological information, for example, inflow, dry/wet condition, and storage state, is considered by reservoir operators with different priorities for the operations of reservoirs with different characteristics (Hejazi et al., 2008). Determining when and under which hydrological condition to switch from one module to another is the key for dynamic operation decisions. The questions to address here are: which hydrological variables are primarily employed in real-world operations of different reservoirs? How is the variable importance related to reservoir characteristics? In order to quantify the importance of an input variable to the release, we use the measure of variable importance in the DT, that is, the feature importance, a built-in output of the CART algorithm, representing the relative importance of an input variable to build a predictive model. For reservoirs with a single module, the variable importance of the single DT is used for those reservoirs. For reservoirs with multi-modules (i.e., multiple DTs), rather than calculating the variable importance for all modules, we just use the variable importance in the classification tree that determines the module application condition. This is because each individual module represents a specific set of release decision rules tailored to a certain hydrological condition, and

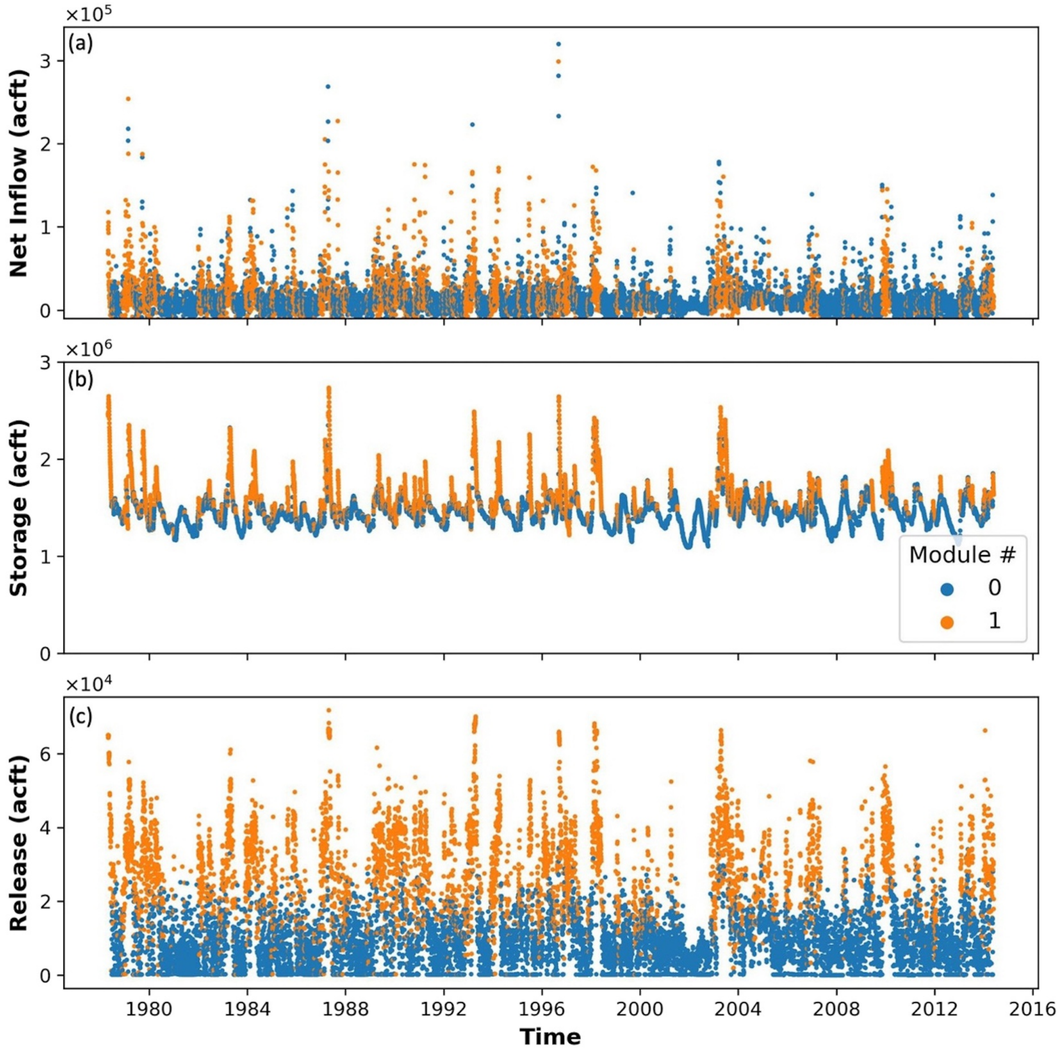

Figure 8. Operation time series of the John H. Kerr Dam in Virginia State: (a) inflow (net), (b) storage, and (c) release. The operation series is labeled by the two operation modules (module 0 and module 1) decoded from GDROM.

the overall importance of a hydrological variable for reservoir operation is prominently captured by modules application conditions characterized by inflow, storage, PDSI, and/or DOY.

Different importance values of input variables are identified for reservoirs with different uses. Figure 9 shows the relative importance of inflow, storage, DOY, and PDSI for reservoirs operated for different primary purposes. For example, the relative importance of inflow is near 1.0 (i.e., the highest importance) for almost all navigation reservoirs, indicating that inflow plays a sole decisive role in operating these reservoirs. Similarly, inflow is recognized as a more important variable than others for reservoirs operated primarily for hydropower because a large portion of the hydropower reservoirs is run-of-river reservoirs. However, for some hydropower reservoirs, other variables such as storage and PDSI also play an important role. It is observed that inflow also plays the most critical role in flood control reservoirs, although the storage impact is non-negligible. As expected, this indicates the importance of high inflows and also reflects the request to maintain a critical storage level for flood control (Ding et al., 2015; Huang et al., 2018). As expected, it is found that DOY is the most important information for irrigation reservoirs, reflecting that the operation of these reservoirs generally follows a seasonal pattern for water refill and release to fulfill agricultural water demands. Finally, reservoirs primarily operated for water supply and recreation are found to have both inflow and storage critically accounted for release decisions, but with slightly higher importance on storage.

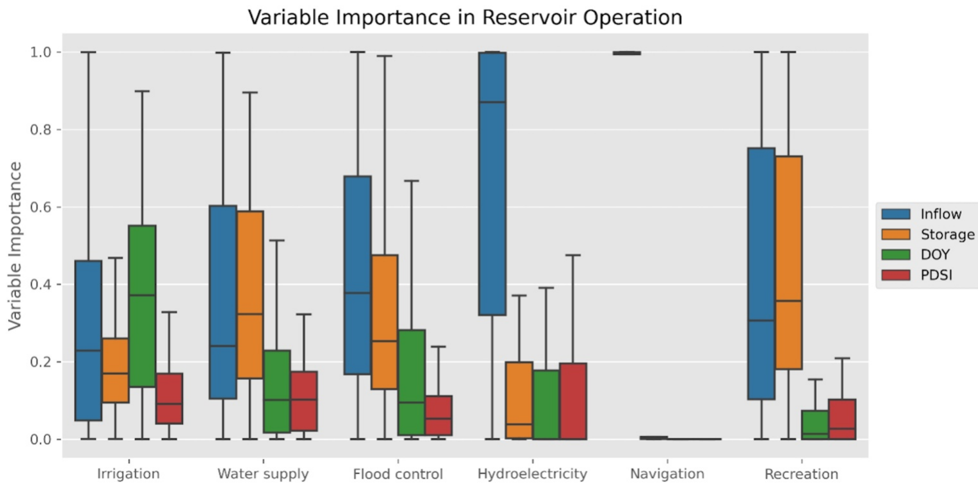

Figure 9. Importance of input variables in reservoir operation, grouped by the primary operation purpose.

The spatial pattern of the role of the various hydrological variables for reservoir operation is further explored. Figures 10a–10c illustrates the spatial distribution of the reservoirs, of which the operations are mainly determined by inflow, DOY, and storage, respectively. Inflow is prevalently adopted as the most critical variable for reservoirs across the CONUS (see Figure 10a), including some flood control reservoirs and irrigation reservoirs with a relatively small regulation capacity, most hydropower reservoirs, and almost all navigation reservoirs. Compared to inflow, DOY is not prevalent at the CONUS level, although DOY is identified as the dominating variable for a large portion of reservoirs in the Western United States, most of which are primarily operated for irrigation (Figure 10b). Reservoirs primarily operated for water supply, recreation, and flood control are clustered in the Southcentral United States, which prioritizes the reservoir storage status over other hydrological information (Figure 10c). Specifically, in Arkansas-White-Red Region, there exist two sets of flood control reservoirs, one of which recognizes inflow as the most critical variable (Figure 10a) and the other prioritizes storage (Figure 10c). This is consistent with the discussion on seasonal operation pattern for flood control reservoirs in Section 3.2, corresponding to two types of flood events in Arkansas-White-Red region: seasonal floods and storm-driven floods. Similar to the irrigation reservoirs, the distinctive use of inflow and storage for the flood control reservoirs is also largely associated with the reservoir regulation capacity.

Finally, it is noted that there are only a small number of reservoirs of which the operation is primarily operated by PDSI. Although many reservoirs play a critical role in mitigating drought stresses, the drought response operation is usually released after experiencing extremely low PDSI for consecutive days or months, which eventually ends with low inflow and dropping storage that a reservoir operator considers more directly. Therefore, PDSI is not commonly found as dominating information in reservoir operations due to the information overlap with inflow and/or storage (see G. Zhao & Gao, 2019 for the relation between PDSI and reservoir surface area and storage).

4. Discussions

4.1. Implications for Improving Reservoir Operation Model Formulation

Previous studies have demonstrated that data analysis findings could be used to improve reservoir operation model formulation, for example, Hejazi and Cai (2011) identified new state variables and used them to formulate a more realistic stochastic dynamic programming model for reservoir operation. Giuliani and Herman (2018) and Q. Zhao and Cai (2020) demonstrated a diagnostic step based on data mining toward a realistic representation of reservoir operations in an optimization model and a simulation model, respectively. Brunner and Naveau (2023) highlighted the potential of using reconstructed seasonal reservoir signals to inform the representation of reservoirs in hydrologic models. Following these studies, the current paper provides greater details of seasonal patterns and discusses how the patterns can be used to inform reservoir model formulation.

As presented above, this study synthesizes the GDROMs for 452 reservoirs in the CONUS to identify typical operation modules (rules) and patterns of module applications and transitions. Furthermore, seasonal patterns of

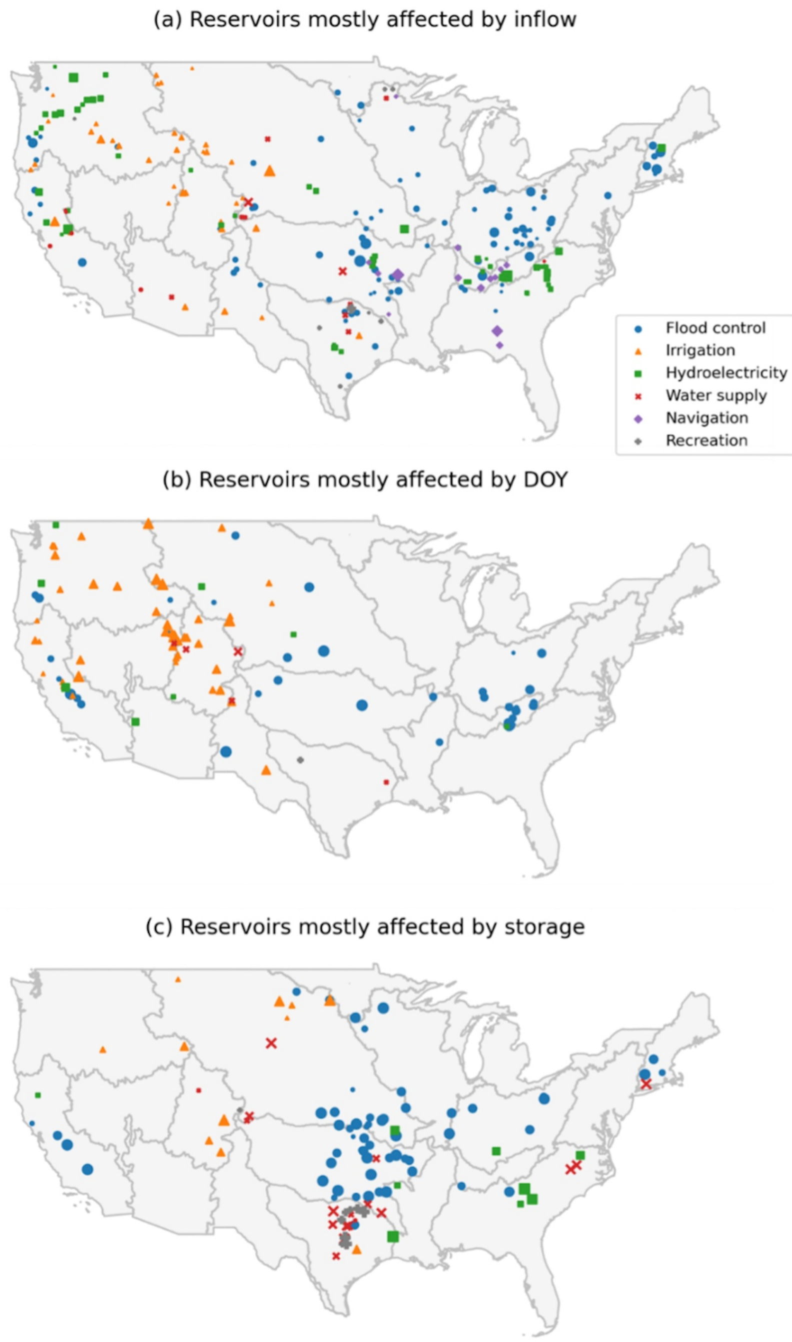

Figure 10. Reservoirs with mostly affected input variable: (a) inflow, (b) DOY, (c) storage. Marker styles and colors distinguish the reservoir primary operation purpose, and marker sizes represent the reservoir regulation capacity, that is, size ratio.

reservoir operation are identified based on the applications and transitions of derived modules within a particular season and across different seasons (Figures 4–7 in Section 3.3). The decoded seasonal information can be built into reservoir operation models to simulate seasonal operation dynamics. Specifically, the five types of modules and the joint module summarized in Table 1 can be used as elementary components to build a new model or improve an existing model developed for a particular reservoir with given operation purpose(s), storage size, climate, and water management institution at a location across the CONUS. In addition, some highlights on seasonal operation patterns are provided below:

- Irrigation reservoirs adopt a typical refill-release cycle characterized by low and constant release mainly applied during the refill period and high release during the high-flow or irrigation season.
- Water supply reservoirs usually adopt a relatively stable release pattern applied most of the year, with high release applied during high-flow or intensive water requirement periods.
- Run-of-river hydropower plants have the inflow-driven release scheme as the principal operation rule; medium to large reservoirs with hydropower as a primary purpose mostly use storage and inflow-driven operation modules jointly.
- Flood control reservoirs use operation rules that are complicated by both climate and flood features. For example, reservoirs in Pacific Northwest Region are operated for flood control during the local rainy season; while those in Ohio Region respond more to winter storms. Additionally, different operation rules are considered for different types of floods. Reservoirs dedicated to episodic and storm-driven floods employ linear release-inflow relationships during the flood days and constantly low release for non-flood days; in contrast, those operated for regular and seasonal floods consider different operation schemes for flood and non-flood seasons and the transition between the two seasons.

On top of the reservoir operation purpose, the design of the operation scheme strongly depends on the reservoir size ratio. In general, inflow-determined modules are formulated for small reservoirs while modules with constant and stable release or storage are adopted for large reservoirs.

It should be noted that in terms of reservoir operation model formulation, the analysis presented in this study highlights the necessity to represent distinct regional features in an operation model (Section 3.3). This request is highlighted by the persistent differences in module application transitions even among reservoirs of similar size and same operation purpose. Most existing models for reservoir operation (e.g., Hanasaki et al., 2006; Yassin et al., 2019; G. Zhao et al., 2016) typically conceptualize the operation transitions based on reservoir storage and seasonal inflow information for a particular operation purpose. Future work will be valuable to simulate the transition between operation modules (or schedules) by considering a complex interaction of multiple factors, including the regional contexts in terms of climate and operation regulations.

4.2. Toward Better Understanding of Standard Operation Policy (SOP) and Hedging Policy (HP)

SOP and HP, the two types of general reservoir operation policies, are extensively studied in the literature (e.g., Draper & Lund, 2004; You & Cai, 2008; Zeng et al., 2021). The module types and module application transition patterns derived from the GDROMs may provide better understanding and modeling of SOP and HP. First, the six typical modules can be referred to different segments of SOP and HP. As shown in Figure S16 in Supporting Information S1, the SOP curve is composed of three linear segments with the following conditions: the release is equal to inflow (i.e., the inflow-driven linear release modules); the release is constant and equal to the full demand before the reservoir is full (i.e., one of constant release module types); the release is equal to the full demand plus the spill when the reservoir is full (i.e., the inflow-driven linear release modules). The HP curve overlaps with the SOP curve by all the segments except for the segment between the hedging start point and endpoint, which can be either linear or nonlinear (Zeng et al., 2021). This special segment of HP is referred to the storage/inflow-driven module type, the storage/inflow-driven piecewise constant release, or approximated by multiple piecewise linear segments (Zeng et al., 2021). The different stages of SOP or HP, which are usually developed for irrigation or water supply reservoirs, respond to intra-year cycle of water refill and release, for example, water conservation via a constant release module and water supply via a linear or nonlinear release module.

As an illustration example of the correspondence between GDROM modules and SOP/HP stages, Figure 11a shows the operation policy of Ochoco Dam in Oregon State, with a primary purpose of irrigation during a selected period (1990–1992), relatively "dry" years in the historical records. The operation of this reservoir is characterized by three modules—constant (during the low flow period), inflow-linear with low inflow levels, and storage-nonlinear when the inflow level reaches a certain level. The inflow-linear module represents the first stage of the SOP (along the 1:1 line, Figure S14 in Supporting Information S1); The storage-nonlinear module represents the nonlinear hedging segment (Figure S16 in Supporting Information S1). Thus, the GDROM of this reservoir shows a typical hedging policy featured by a two-phase release policy: an initial linear phase to meet current demand (a common stage for both SOP and HP) and a follow-up nonlinear phase under HP. In addition, Figure 11b demonstrates the operation of the Eucha Dam in Oklahoma which is primarily used for water supply. There is a clear trend of near-constant releases when WA falls below a certain threshold (approximately 76,000

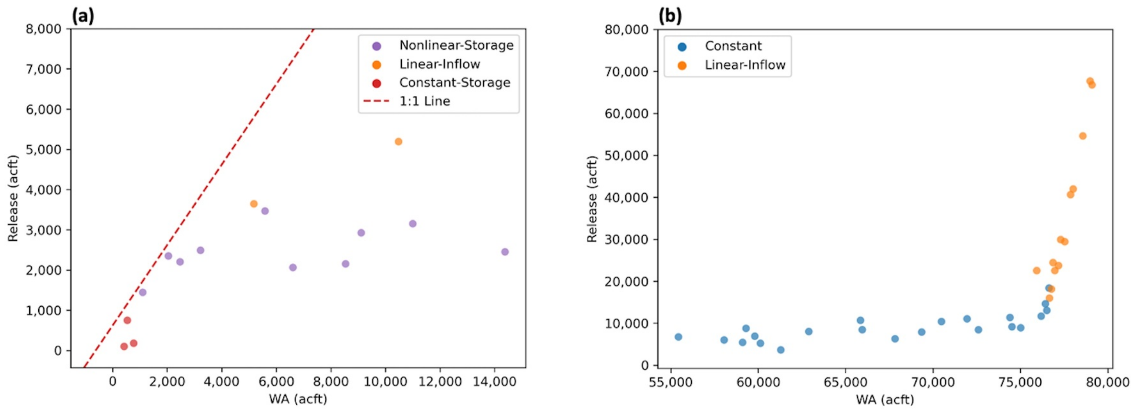

Figure 11. Examples of real-world operation demonstrating the segments of SOP/HP. (a) Operation of the Ochoco Dam in Oregon State during 1990–1992, a dry period. Low-release months during winter are not displayed in order to highlight the hedging operation part during major supply months. (b) Operation of the Eucha Dam in Oklahoma state during 2009–2011. Each point shows the pair of monthly water availability (WA) and release. Release and WA are aggregated into monthly values.

acre-feet), beyond which release increases with WA linearly. The near-constant releases, fit the flat stage of the SOP shown in Figure S14 in Supporting Information S1; the release linearly increasing with WA fits the spill stage when WA discounted by the constant release is larger than the maximum storage (a segment for both SOP and HP curve shown in Figure S14 in Supporting Information S1).

Moreover, in existing SOP and HP studies, the "water availability" item aggregates inflow and storage, thus the condition that justifies the transition of the various operation stages (Figure S14 in Supporting Information S1) does not specify the separate roles of inflow and storage regarding their contribution to water availability. The six operation module types separate the rules by storage and release, which reflects more specific conditions of water availability and can provide more operation options, especially with HP, for example, strategic control of storage, higher priority for inflow forecast considering uncertainty and risk aversion, etc. Thus, the six types of modules and application transition patterns can be used to improve SOP or HP with more specified inputs, that is, inflow, storage, and/or both.

5. Summary and Conclusions

This study synthesizes the outputs of generic data-driven reservoir operation models (GDROMs; Chen et al., 2022) developed for 452 data-rich reservoirs of different sizes, operation purposes, and locations (climates and regulations) across the CONUS. Based on the synthesis, typical operation modules and their application transition patterns are uncovered, and major driving forces are identified for the release decisions.

It is found that a small number of modules (up to eight in our findings) can effectively represent the operation of any of those reservoirs in sample. Most hydropower and navigation reservoirs use a single module, while flood-control and irrigation reservoirs mostly use multiple modules, reflecting the adoption of multiple operation schemes and seasonal transitions in operating these reservoirs. Furthermore, the operation modules are categorized into five basic types and joint modules (summarized in Table 1). Notably, constant release modules are common in reservoirs with large regulation capacity; irrigation reservoir operation can generally be modeled using low-release and high-release modules and their intra-year transitions, with regional variability adjusted. In contrast, the operation of flood control reservoirs is modeled with a more complex structure, with consideration of regional climate characteristics and flood types (storm-driven or seasonal). The roles of each hydroclimatic variable (inflow, storage, PDSI, and DOY) are prioritized for release decisions of reservoirs operated for different purposes. It is found that for irrigation reservoirs in the western United States, the role of inflow and DOY depend on the reservoir size ratio (ratio of storage capacity to annual inflow), with inflow dominating small reservoirs and DOY dominating large ones. For flood control reservoirs in the central US, inflow is mostly considered by reservoirs with a relatively small storage capacity and storage is mostly considered for those with a large storage capacity. Similarly, for hydropower reservoirs, inflow is prioritized information for relatively small reservoirs (when the size ratio is less than 0.3); while storage is mostly found as the prioritized information for relatively large reservoirs.

The discovered typical modules and their application transition patterns extend the rules and conditions of standard operation policy (SOP) and hedging policy (HP), and thus can be informative for building more realistic SOP and HP models by using respective inflow and storage conditions (i.e., the rules corresponding to different segments of SOP and HP can be conditioned with either inflow or storage, or both). Moreover, the findings with the 452 reservoirs across the CONUS can be used to build a data-driven operation model for reservoirs with limited data availability, especially those in developing countries. In this way, the GDROMs can be extended to data-scarce reservoirs. Furthermore, since GDROMs have an interpretable and transparent decision tree-based model structure and four input variables, they can be easily incorporated into the rainfall-runoff processes of a large-scale hydrologic model, as a generic reservoir modeling module. Thus, GDROMs can be interesting to the community of large-scale hydrologic modeling (Vora et al., 2024).

We have to admit that the derived five basic types of operation modules from 452 CONUS reservoirs may not be applicable to some other reservoirs. The patterns can be validated and applied to new reservoirs when more data are available. Moreover, it should be noted that the data-driven models (GDROMs) used in this work can be limited in representing flood control rules given that some reservoirs are operated in sub-daily time intervals during a flood event. In addition, the GDROM performs relatively weak (Chen et al., 2022) with abnormally high inflows and releases in some regions such as the North and Mid-Atlantic. In the future, the GDROMs can be improved by using robust loss functions to overcome the bias caused by extreme values (outliers) to replace the currently employed mean square error (Hastie et al., 2009b). Alternatively, the model structure may be modified to fit the extreme and regular releases separately, that is, the model could be specifically applied and tuned for flood events for improved simulation of extremely high releases.

Data Availability Statement

The historical operation data is retrieved from the USBR water operation database (USBR, 2021), accessed separately for each USBR Region via the "Water Operations" link; the WM data dissemination (USACE, 2021) via the "Project Webservices" section under "Data Discovery," and the ResOpsUS database (Steyaert et al., 2022b). The inventory of trained operation rules and the training data are available at the HydroShare repository established by Li et al. (2023). The code and notebook for building and running the GDROM is available at the Zenodo repository (Li & Chen, 2023).

Acknowledgments
The authors are grateful for multiple reviews from the Associate Editor and two reviewers, which have led to significant improvement of this paper. We also thank Kevin Wallington for his editorial help. Funding for this project was provided by the National Oceanic and Atmospheric Administration (NOAA), awarded to the Cooperative Institute for Research on Hydrology (CIROH) through the NOAA Cooperative Agreement with The University of Alabama, NA22NWS4320003 (A23-0242-S001-A01 "CIROH: Integrated Streamflow and Reservoir Release Forecast to Mitigate Drought Effects at the River Basin Scale").

References

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, *52*(6), 4209–4225. https://doi.org/10.1002/2015WR017864

Bessler, F. T., Savic, D. A., & Walters, G. A. (2003). Water reservoir control with data mining. *Journal of Water Resources Planning and Management*, *129*(1), 26–34. https://doi.org/10.1061/(asce)0733-9496(2003)129:1(26)

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees. Statistics/probability series*. Wadsworth & Brooks/Cole Advanced Books & Software.

Brunner, M. I., & Naveau, P. (2023). Spatial variability in Alpine reservoir regulation: Deriving reservoir operations from streamflow using generalized additive models. *Hydrology and Earth System Sciences*, *27*(3), 673–687. https://doi.org/10.5194/hess-27-673-2023

Chen, Y., Li, D., Zhao, Q., & Cai, X. (2022). Developing a generic data-driven reservoir operation model. *Advances in Water Resources*, *167*, 104274. https://doi.org/10.1016/j.advwatres.2022.104274

Coerver, H. M., Rutten, M. M., & Van De Giesen, N. C. (2018). Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences*, *22*(1), 831–851. https://doi.org/10.5194/hess-22-831-2018

Conway, D., & Mahé, G. (2009). River flow modelling in two large river basins with non-stationary behaviour: The Paraná and the Niger. *Hydrological Processes*, *23*(22), 3186–3192. https://doi.org/10.1002/hyp.7393

Corani, G., Rizzoli, A. E., Salvetti, A. N. D. R. E. A., & Zaffalon, M. (2009). Reproducing human decisions in reservoir management: The case of lake Lugano. In *Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium Thessaloniki, Greece, May 28–29, 2009* (pp. 252–263). Springer Berlin Heidelberg.

Deng, C., Liu, P., Guo, S., Wang, H., & Wang, D. (2015). Estimation of nonfluctuating reservoir inflow from water level observations using methods based on flow continuity. *Journal of Hydrology*, *529*, 1198–1210. https://doi.org/10.1016/j.jhydrol.2015.09.037

Ding, W., Zhang, C., Peng, Y., Zeng, R., Zhou, H., & Cai, X. (2015). An analytical framework for flood water conservation considering forecast uncertainty and acceptable risk. *Water Resources Research*, *51*(6), 4702–4726. https://doi.org/10.1002/2015WR017127

Draper, A. J., & Lund, J. R. (2004). Optimal hedging and carryover storage value. *Journal of Water Resources Planning and Management*, *130*(1), 83–87. https://doi.org/10.1061/(asce)0733-9496(2004)130:1(83)

Ehsani, N., Fekete, B. M., Vörösmarty, C. J., & Tessler, Z. D. (2016). A neural network based general reservoir operation scheme. *Stochastic Environmental Research and Risk Assessment*, *30*(4), 1151–1166. https://doi.org/10.1007/s00477-015-1147-9

Giuliani, M., & Herman, J. D. (2018). Modeling the behavior of water reservoir operators via eigen behavior analysis. *Advances in Water Resources*, *122*, 228–237. https://doi.org/10.1016/j.advwatres.2018.10.021

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. *Water Resources Research*, *35*(4), 1305–1311. https://doi.org/10.1029/1999WR900016

Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, *327*(1), 22–41. https://doi.org/10.1016/j.jhydrol.2005.11.011

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009a). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed., pp. 219–254). springer.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009b). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed., pp. 346–350). springer.

Hejazi, M. I., & Cai, X. (2009). Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. *Advances in Water Resources*, *32*(4), 582–593. https://doi.org/10.1016/j.advwatres.2009.01.009

Hejazi, M. I., & Cai, X. (2011). Building more realistic reservoir optimization models using data mining–A case study of Shelbyville Reservoir. *Advances in Water Resources*, *34*(6), 701–717. https://doi.org/10.1016/j.advwatres.2011.03.001

Hejazi, M. I., Cai, X., & Ruddell, B. L. (2008). The role of hydrologic information in reservoir operation—Learning from historical releases. *Advances in Water Resources*, *31*(12), 1636–1650. https://doi.org/10.1016/j.advwatres.2008.07.013

Huang, K., Ye, L., Chen, L., Wang, Q., Dai, L., Zhou, J., et al. (2018). Risk analysis of flood control reservoir operation considering multiple uncertainties. *Journal of Hydrology*, *565*, 672–684. https://doi.org/10.1016/j.jhydrol.2018.08.040

Labadie, J. W. (2004). Optimal operation of multireservoir systems: State-of-the-art review. *Journal of Water Resources Planning and Management*, *130*(2), 93–111. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(93)

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Li, D., & Chen, Y. (2023). lidh966/GDROM: Generic data-driven reservoir operation model (v1.0.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.8018970

Li, D., Chen, Y., Cai, X., & Zhao, Q. (2023). Data-driven reservoir operation rules for 450+ reservoirs in contiguous United States [Dataset]. *HydroShare*. https://doi.org/10.4211/hs.63add4d5826a4b21a6546c571bdece10

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Oliveira, R., & Loucks, D. P. (1997). Operating rules for multireservoir systems. *Water Resources Research*, *33*(4), 839–852. https://doi.org/10.1029/96wr03745

Palmer, W. C. (1965). In *Meteorological drought* (Vol. 30). US Department of Commerce.

Steyaert, J. C., Condon, L. E., Turner, W. D., & Voisin, N. (2022b). ResOpsUS [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.5367383

Steyaert, J. C., Condon, L. E., Turner, W. D., & Voisin, N. (2022a). ResOpsUS, a dataset of historical reservoir operations in the contiguous United States. *Scientific Data*, *9*(1), 34. https://doi.org/10.1038/s41597-022-01134-7

Turner, S. W. D., Steyaert, J. C., Condon, L., & Voisin, N. (2021). Water storage and release policies for all large reservoirs of conterminous United States. *Journal of Hydrology*, *603*, 126843. https://doi.org/10.1016/j.jhydrol.2021.126843

Turner, S. W. D., Xu, W., & Voisin, N. (2020). Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrology and Earth System Sciences*, *24*(3), 1275–1291. https://doi.org/10.5194/hess-24-1275-2020

USACE. (1995). Project manual for water control management Senecaville Lake—Seneca fork. Retrieved from https://water.usace.army.mil/a2w/CWMS_CRREL.cwms_util_api.download_dcp?p_dcp_document_id=2223

USACE. (2021). WM data dissemination [Dataset]. Retrieved from https://water.usace.army.mil/

USBR. (2021). Bureau of reclamation water operations [Dataset]. Retrieved from https://www.usbr.gov/main/water/

Vora, A., Cai, X., Chen, Y., & Li, D. (2024). Coupling reservoir operation and rainfall-runoff processes for streamflow simulation in watersheds. *Water Resources Research*, *60*(6), e2023WR035703. https://doi.org/10.1029/2023WR035703

Wallington, K., & Cai, X. (2020). Feedback between reservoir operation and floodplain development: Implications for reservoir benefits and beneficiaries. *Water Resources Research*, *56*(4), e24524. https://doi.org/10.1029/2019wr026610

Wurbs, R. A. (1993). Reservoir-system simulation and optimization models. *Journal of Water Resources Planning and Management*, *119*(4), 455–472. https://doi.org/10.1061/(asce)0733-9496(1993)119:4(455)

Yang, S., Yang, D., Chen, J., & Zhao, B. (2019). Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *Journal of Hydrology*, *579*, 124229. https://doi.org/10.1016/j.jhydrol.2019.124229

Yang, T., Gao, X., Sorooshian, S., & Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resources Research*, *52*(3), 1626–1651. https://doi.org/10.1002/2015WR017394

Yang, T., Liu, X., Wang, L., Bai, P., & Li, J. (2020). Simulating hydropower discharge using multiple decision tree methods and a dynamical model merging technique. *Journal of Water Resources Planning and Management*, *146*(2), 04019072. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001146

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., & Wheater, H. (2019). Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences*, *23*(9), 3735–3764. https://doi.org/10.5194/hess-23-3735-2019

Yeh, W. W.-G. (1985). Reservoir management and operations models: A state-of-the-art review. *Water Resources Research*, *21*(12), 1797–1818. https://doi.org/10.1029/WR021i012p01797

You, J.-Y., & Cai, X. (2008). Hedging rule for reservoir operations: 1. A theoretical analysis. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006WR005481

Zeng, X., Lund, J. R., & Cai, X. (2021). Linear versus nonlinear (convex and concave) hedging rules for reservoir optimization operation. *Water Resources Research*, *57*(12), e2020WR029160. https://doi.org/10.1029/2020WR029160

Zhao, G., & Gao, H. (2019). Towards global hydrological drought monitoring using remotely sensed reservoir surface area. *Geophysical Research Letters*, *46*(13), 13027–13035. https://doi.org/10.1029/2019GL085345

Zhao, G., Gao, H., Naz, B. S., Kao, S. C., & Voisin, N. (2016). Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources*, *98*, 16–31. https://doi.org/10.1016/j.advwatres.2016.10.014

Zhao, Q., & Cai, X. (2020). Deriving representative reservoir operation rules using a hidden Markov-decision tree model. *Advances in Water Resources*, *146*, 103753. https://doi.org/10.1016/j.advwatres.2020.103753